Motion Imagery Standards Board **Recommended Practice** **Time Stamping Compressed Motion Imagery**	**MISB RP 0604** **Date 13 June 2007**

1 Scope

This Recommended Practice defines methods to time stamp compressed video streams and to synchronously transport video and metadata in compressed motion imagery streams.

Implementation methods are defined that leverage the transport layer of MPEG-2 for carriage of motion imagery streams of varying types and bit rates as defined in the Motion Imagery Standards Profile concept of X on 2. Specific formats covered include MPEG-2 and H.264.

2 References

ISO/IEC 13818-1:2000 Information technology — Generic coding of moving pictures and associated audio information: Systems

ISO/IEC 13818-1:2000/Amd 1:2003 Information technology — Generic coding of moving pictures and associated audio information: Systems, AMENDMENT 1: Carriage of metadata over ISO/IEC 13818-1 streams

ISO/IEC 13818-1:2000/Amd 3:2004 - Transport of AVC video data over ITU-T Rec H.222.0 | ISO/IEC 13818-1 streams

ISO/IEC 13818-2 Information technology — generic coding of moving pictures and associated audio information: Video

ISO/IEC 14496-10 Information technology — coding of audio-visual objects — Part 10: Advanced video coding

MISB RP 0101 Use of MPEG-2 System Streams in Digital Motion Imagery Systems

MISB RP 0603, Common Time Reference for Digital Motion Imagery using Coordinated Universal Time (UTC)

Motion Imagery Standards Profile 3.5, DoD/IC Motion Imagery Standards Board, http://www.ismc.nga.mil

SMPTE 12M-1999, Television, Audio and Film - Time and Control Code

SMPTE RP 188 – 1999, Transmission of Time Code and Control Code in the Ancillary Data Space if a Digital Television Data Stream

SMPTE 328M-2000, MPEG-2 Video Elementary Stream Editing Information

SMPTE RP 217-2001, Nonsynchronized Mapping of KLV Packets into MPEG-2 System Streams

SMPTE 309M-1999, Transmission of Date and Time Zone Information in Binary Groups of Time and Control Code

SMPTE 336M-2001 Data Encoding Protocol using Key-Length-Value

3 Introduction

The MPEG-2 transport layer (ISO/IEC 13818-1:2000) provides an infrastructure for the carriage of video, audio and metadata in a single motion imagery stream as shown in the following diagram.

Figure 1: MPEG-2 Transport Stream

The Motion Imagery Standards Profile (MISP) endorses the use of MPEG-2 Transport Streams for this purpose. The MISB has been researching the use of MPEG-2 Transport Streams for the carriage of other motion imagery formats in a study known as X on 2. Recent recommendations extend the use of MPEG-2 Transport Streams as a means for carriage of H.264 video in the compressed domain as defined in ISO/IEC 13818-1 Amd3:2004.

MISB RP 0603 outlines the advantages of using Universal Coordinated Time (UTC) as the master clock reference for video and metadata, and discusses several time formats and the relationships between them. This RP defines how the UTC time can be used to stamp MPEG-2 and H.264 video streams, and how the video and metadata can be synchronously transported in motion imagery streams.

Motion imagery analysis and processing applications require various levels of temporal accuracy when referencing metadata elements and the video frames associated with those elements. Compressed imagery generated from standard definition analog video sensors has traditionally utilized asynchronous methods for carriage of metadata in a private data stream. This was adequate for metadata that was not time sensitive, or metadata which only needed to be associated to within a few seconds of the correct video frame. Asynchronous transport could not be used reliably for systems which required metadata to be frame or sub-frame accurate.

Synchronous multiplexing of metadata with video ensures that the proximity between a metadata item and the associated video is well defined. This in turn reduces the latency in the system and helps prevent the metadata from being separated from the associated video when the video is processed.

This RP shall provide guidance on methods to synchronously transport video frames and associated metadata elements with varying levels of precision as determined by the user's requirements.

4 Time Stamping Video

System designers should be aware of the accuracy requirements for the time stamps in their system. RP 0603 defines the use of UTC as a deterministic common time reference for the correlation of motion imagery frames and metadata. It also describes several types of systems and the relative accuracies of each.

Time stamps may be introduced into a compressed video stream in one of two ways. If the uncompressed video signal contains a time stamp in the Vertical Interval Time Code (VITC) or the Vertical Ancillary Data Space (VANC), it is recommended that encoder extract the time stamp from the VITC or VANC of the incoming video signal and insert it into the video elementary stream as indicated in the following sections.

If the uncompressed video signal does not contain a time stamp, the encoder should be set up to read the time stamp from the system time clock or an external source and insert it into the video elementary stream.

The following sections describe how to insert the time stamp into MPEG-2 and H.264 video streams.

4.1 MPEG-2

MISP Standard 9715 "Time Reference Synchronization" states that:

> Universal coordinated time (UTC, also known as "Zulu"), clock signals shall be used as the universal time reference for DoD/IC/NSG SMPTE 12M time code systems, allowing systems using time code to accurately depict the actual Zulu time of day of motion imagery acquisition / collection / operations.

The following sections describe how to use the Group of Pictures (GOP) time code to time stamp MPEG-2 compressed video, and how a time stamp in the video elementary stream user data field or a time stamp in the MPEG-2 video elementary stream editing information may be used in systems which require a more persistent time stamp or one with a higher level of precision.

4.1.1 GOP Time Code

The MPEG-2 video layer includes the definition of a time code within the GOP header. This time code is of the form HH:MM:SS:FF in a format specified by SMPTE 12M-1999.

It is strongly recommended that the SMPTE time code in the GOP header be filled in with a time stamp which represents UTC time for MPEG-2 video streams for all motion imagery systems.

MISB RP 0603 gives an indication of the accuracy of the SMPTE 12M time code as it is inserted into the video signal for systems with integer and non-integer frame rates, and for cameras which are or are not phase locked to the master time reference.

For systems which process signals with integer frame rates, and for video sources that are genlocked to a UTC time reference, the accuracy of the time stamp in the GOP header can be quite accurate (sub-frame accuracy). The accuracy decreases for systems with non-integer frame rates.

The usefulness of the GOP time code has some limitations:

- The GOP time code is generally not persistent, and not considered by the MPEG-2 standards to be an absolute time. When a video is edited, the editor will often re-stamp the GOP time code.
- The GOP time code includes a time, but not a date. The date information, if needed, must be extracted from the KLV metadata in the stream.
- The accuracy of the GOP time code is limited, particularly in motion imagery with non-integer frame rates.

Some of these limitations can be addressed by also populating a time code in the elementary stream user data or MPEG-2 video elementary stream editing information as described in the following sections.

4.1.2 Elementary Stream User Data

The MPEG-2 format allows user defined data to be inserted into the video elementary stream in a user data field (start code of 0xB2). The ISO/IEC 13818-2 specification allows the user data field to be placed in several different places in the video bitstream. The user data field containing the time stamp is placed between the picture header and the picture data so that it relates to a frame of video.

The elementary stream user data time stamp may be used in systems which are required to associate a highly accurate, microsecond resolution time stamp with the video frame. This UTC time stamp shall be derived from GPS as described in section 4 of RP 0603 and will be formatted as defined in Annex A of RP 0603.
The user data message consists of an identification string and a time stamp as defined below:

Identification String: 16 bytes that shall be set to the value:

Bytes 1-8: 0x4D, 0x49, 0x53, 0x50, 0x6D, 0x69, 0x63, 0x72,
Bytes 9-16: 0x6F, 0x73, 0x65, 0x63, 0x74, 0x69, 0x6D, 0x65

This represents the ASCII sting: “MISPmicrosectime”

Time Stamp: 12 additional bytes defined as follows:

Byte 17: Status
- Bit 7 0=GPS Locked (internal clock locked to GPS)
  1=GPS Flywheel (internal clock not locked to GPS, so it is running on an internal oscillator)
- Bit 6 0=Normal (time incremented normally since last message)
  1=Discontinuity (time has not incremented normally since last message)
- Bit 5 0=Forward (If Bit 6=1, this indicates that the time jumped forward)
  1=Reverse (If Bit 6=1, this indicates that the time jumped backwards)
- Bits 4-0: Reserved (=1)

Bytes 18, 19: Two MS bytes of Time Stamp (microseconds)
Byte 20: Start Code Emulation Prevention Byte (0xFF)
Bytes 21, 22: Two next MS bytes of Time Stamp (microseconds)
Byte 23: Start Code Emulation Prevention Byte (0xFF)
Bytes 24, 25: Two next LS bytes of Time Stamp (microseconds)
Byte 26: Start Code Emulation Prevention Byte (0xFF)
Bytes 27, 28: Two LS bytes of Time Stamp (microseconds)

This represents the 64 bit microsecond UTC time stamp where byte 18=MSB, bytes 19,21,22,24,25,27 are intermediate bytes and byte 28=LSB. Byte 1 is transmitted first.

4.1.3 MPEG-2 Video Elementary Stream Editing Information

SMPTE 328M describes additional information that may be carried in the user data area of a video elementary stream. One of the additional metadata elements is a 64 bit time code which complies with SMPTE 12M:1999 and SMPTE 309M:1999. The time code represents the time that the frame was captured (HH:MM:SS:FF), and it contains a date as defined in SMPTE 309M.

4.2 H.264

As with MPEG-2, the H.264 compression format provides places to include a time stamp in the video stream. Both of the time stamps described below are placed in the Supplemental Enhancement Information (SEI) message.

4.2.1 Pic_Timing Time Stamp

The H.264 format, specified in ISO/IEC14496-10 provides for an optional time stamp to be defined in the SEI Message. The “picture timing SEI message” (pic_timing) specifies the time as HH:MM:SS:FF. It is a persistent time stamp which reflects the time of frame

capture and it contains flags to specify whether the video is drop-frame, and whether there is a discontinuity in the video time-line.

For H.264 compressed motion imagery, it is strongly recommended that the pic_timing field in the SEI Message be filled in with a time stamp which represents UTC time for H.264 video streams for all motion imagery systems.

4.2.2 User Data

The H.264 format also allows user defined data to be associated with a particular video frame using the user data unregistered SEI Message.

The user data unregistered SEI Message may be used in systems which are required to associate a highly accurate, microsecond resolution time stamp with the video frame. This UTC time stamp shall be derived from GPS as described in section 4 of RP 0603 and will be formatted as defined in Annex A of RP 0603.

The user data unregistered message consists of two fields as defined below:
Uuid_iso_iec_11578 is a 16 byte field that shall be set to the value:

Bytes 1-8:	0x4D, 0x49, 0x53, 0x50, 0x6D, 0x69, 0x63, 0x72,
Bytes 9-16:	0x6F, 0x73, 0x65, 0x63, 0x74, 0x69, 0x6D, 0x65

This represents the ASCII sting: “MISPmicrosectime”

User_data_payload_bytes is a variable length field. For this application, 12 bytes will be used as follows:

Byte 1: Status
- Bit 7 0=GPS Locked (internal clock locked to GPS)
  1=GPS Flywheel (internal clock not locked to GPS, so it is running on an internal oscillator)
- Bit 6 0=Normal (time incremented normally since last message)
  1=Discontinuity (time has not incremented normally since last message)
- Bit 5 0=Forward (If Bit 6=1, this indicates that the time jumped forward)
  1=Reverse (If Bit 6=1, this indicates that the time jumped backwards)
- Bits 4-0: Reserved (=1)

Bytes 2, 3: Two MS bytes of Time Stamp (microseconds)
Byte 4: Start Code Emulation Prevention Byte (0xFF)
Bytes 5, 6: Two next MS bytes of Time Stamp (microseconds)
Byte 7: Start Code Emulation Prevention Byte (0xFF)
Bytes 8, 9: Two next LS bytes of Time Stamp (microseconds)
Byte 10: Start Code Emulation Prevention Byte (0xFF)
Bytes 11, 12: Two LS bytes of Time Stamp (microseconds)

This represents the 64 bit microsecond UTC time where byte 2=MSB, bytes 3,5,6,8,9,11 are intermediate bytes and byte 12=LSB. Byte 1 is transmitted first.

5 Time Stamping Metadata

Systems which are capable of time stamping both the video stream and the metadata stream have all of the information necessary to multiplex this information together in a synchronized motion imagery stream.

SMPTE 336M defines the structure of KLV metadata. The KLV element "User Defined Time Stamp (microseconds since 1970)" is typically used as the time stamp in a KLV stream. The definition and format of this KLV element are defined in RP 0603.

6 Carriage of Metadata in Transport Stream

If the requirements for a motion imagery system dictate that a metadata element is associated with a particular frame of video, or that the time associated with the metadata element is correlated to the same time line as the video, then ISO/IEC 13818-1/Amd1 shall be used to transport the video and associated metadata in an MPEG-2 Transport Stream.

6.1 Asynchronous Carriage of Metadata (RP217)

SMPTE RP 217 is currently used to define the transport of KLV metadata over MPEG-2 transport streams in an asynchronous way. The metadata packets do not have Presentation Time Stamps (PTS). The relationship between the metadata and the video frames is typically established by their proximity in the video stream. This type of metadata carriage may be used to transport static metadata, or metadata which is not tied closely in time to the video.

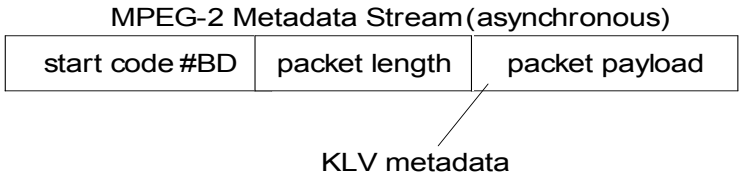

Figure 2: Asynchronous Metadata Stream

6.2 Synchronous Carriage of Metadata

ISO/IEC 13818-1 Amd1 details several ways to carry metadata over MPEG-2 transport streams. This RP recommends the method outlined in Section 2.12.4 "Use of PES packets to transport metadata" for transporting metadata that is synchronized with the video stream. This method provides a way to synchronize metadata with video using the Presentation Time Stamp (PTS) found in the Packetized Elementary Stream (PES) header. This time stamp is in the MPEG-2 systems layer, and is relevant for H.264 as well as MPEG-2.

The metadata may or may not be sampled at the same time as a video frame depending upon the system design. If it is sampled at the same time as a video frame,

the metadata and video frame will have the same PTS. If the metadata is not sampled at the same time as the video frame, it will be stamped with a different PTS, but exist on the same timeline as the video frame.

Figure 3 shows the general structure of a PES packet in the metadata bit stream. In the most common implementation, the packet payload would consist of a single metadata cell which includes a five byte header, and KLV metadata.

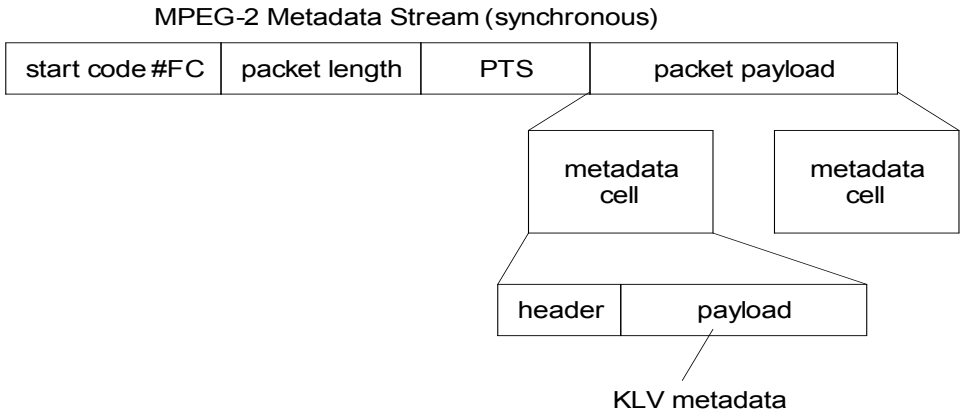

Figure 3: Synchronous Metadata Stream

A metadata service is defined in ISO/IEC 13818-1 Amd 1 as "a coherent set of metadata of the same format delivered to a receiver for a specific purpose." When transporting metadata using an Amd 1 stream, a unique metadata service id is assigned to each service. Each metadata service is represented by a collection of metadata access units which are transported in PES packets.

Details of the implementation of this method are given below. These recommendations were extracted from ISO/IEC 13818-1/Amd1, and ISO/IEC 13818-1:2000.

Metadata Stream

- A "descriptive data stream" (Stream Id = 0xFC) shall be used.
- The Stream Type for each PES packet shall be 0x15 "Metadata carried in PES packets using the Metadata Access Unit Wrapper"
- Each PES packet shall have a PTS to be used to synchronize the metadata with the video frames.
- In each PES packet that carries metadata, the first PES packet data byte shall be the first byte of a Metadata Access Unit Cell.

- The PTS in the PES header shall apply to each Access Unit contained in the PES packet.

- The PTS shall signal the time that the metadata Access Unit becomes relevant. It is assumed that the metadata is decoded instantaneously (no DTS needed). If a video frame and a metadata access unit have the same PTS, then they were sampled at the same time. .

- Each metadata Access Unit may be carried in one or more Access Unit Cells.

- (from 13818-1:2000) "The delay of any data through the System Target Decoder buffers shall be less than or equal to one second" (This ensures that the metadata is close in the stream to the video frames that it relates to.)

  Note: Careful use of the buffer size and leak rate for metadata defined in the System Target Decoder (STD) Model (and specified in the Metadata STD Descriptor) can force a closer proximity of the metadata to the associated frame of video.

Program Map Table (PMT)

- The Metadata Stream shall be defined in the PMT as a separate Stream within the same Program as the Video Elementary Stream. 13818-1 Amd1 allows for multi-program Transport Streams, and methods for associating metadata in one program to video in another. Multi-program Transport Streams are not covered within the scope of this RP.

The PMT shall contain a Metadata Descriptor for each metadata service within the metadata stream. The Metadata Descriptor(s) shall be within the descriptor loop for the metadata stream. The Metadata Descriptor contains the metadata_service_id for the service it describes. The following values are used to identify metadata types within the Metadata Descriptor:

metadata_format = 0xff (specified by metadata format identifier)
metadata_format_identifier = "KLVA" (for all types of KLV metadata)
DSM-CC_flag = 0
decoder_config_flags = 000

Metadata_application_format (type of KLV metadata)	
0x0100	General
0x0101	Geographic Metadata
0x0102	Annotation Metadata

0x0103	Still Image on Demand

Metadata_application_format is responsible for defining the syntax and meaning of the service_identification_record and private data bytes within the Metadata Descriptor. For all of the above metadata_application_format values, service_identification_record and private data bytes are not allowed. Note that this implies DSM-CC_flag is set to zero.

NOTE: These metadata_application_format values are within a range specified as "User Defined" in ISO/IEC 13818-1 Amd 1. The definitions above are only valid for streams generated by systems compliant with this document and, thus, should be considered as informational rather than authoritative. Parsers should not depend on these values having the provided meanings for streams whose origins are not known to originate from systems compliant with this document.

The following is a sample Metadata Descriptor for a metadata stream containing geographic KLV metadata.

Metadata Descriptor Content
descriptor tag = 0x26
descriptor_length = 0x09
metadata_application_format = 0x0101
metadata_format = 0xff
metadata format identifier = 0x4b 0x4c 0x56 0x41 = "KLVA"
metadata_service_id = 0x00
decoder_config_flags, DSM-CC_flag, reserved = 0x00

- The PMT shall contain a single Metadata STD Descriptor for the metadata stream.

- The PMT may contain other descriptors such as the "Content Labeling Descriptor" and the "Metadata Pointer Descriptor".

7 Transition Annex

Many motion imagery systems have been developed based on SMPTE RP217 for asynchronous carriage of metadata in MPEG-2 Transport Streams. ISO/IEC 13818-1/Amd1 provides a synchronous method for transporting metadata with the associated video streams. As systems advance and strive for more accurate metadata, migrating to this new method of transporting metadata is important.

New systems and applications must be capable of handling metadata using both the SMPTE RP217 format and the ISO/IEC 13818-1/Amd1 format. It will be relatively straight forward for motion imagery systems to add support for ISO/IEC 13818-1/Amd1. Minor changes must be made to the transport layer (multiplexing and demultiplexing) of the motion imagery stream. During the transition period, the MISB will provide tools to

convert data from the Amd 1 format into the legacy SMPTE RP217 format. Some synchronization information will be lost in this conversion, but it will allow legacy software to be able to interpret the video and metadata.